BE/APh 161: Physical Biology of the Cell, Winter 2014
Homework #4
Due at the start of lecture, 1PM, February 12, 2014.

Problem 4.1 (Comments on *Cell Biology by the Numbers* part 4, 10 pts)**.**
We continue in our reading in *CBBTN*. This time, please read chapter 3, pages 151–199, and send comments about two vignettes. Remember to email your answers to me and the TAs and indicate whether you would like to be anonymous when I send the comments to the book's authors. Also, please either send your responses as text in an email or as a PDF. Do not send MS Word documents.

Problem 4.2 (Information and the MWC model, 15 pts)**.**
The Marzen, et al. paper in your reading presented an information-theoretic perspective of the MWC concept. We will explore some of the key concepts presented there in this problem. In particular, we consider a cell containing N ion channels as described in lecture (and also in section 7.3 of *PBoC2* and in Figure 3 of Marzen, et al.).

a) Explain why the probability that N_{open} of the ion channels are open, given a ligand concentration of c, is given by

$$P(N_{\text{open}}|c) = \frac{N!}{(N - N_{\text{open}})! N_{\text{open}}!} (P_{\text{open}}(c))^{N_{\text{open}}} (1 - P_{\text{open}}(c))^{N - N_{\text{open}}}. \tag{4.1}$$

b) The conditional entropy of a random variable X given a random variable Y is

$$H(X|Y) = \sum_y P(y) \left(- \sum_x P(x|y) \log_2 P(x|y) \right) = - \sum_x \sum_y P(y) P(x|y) \log_2 P(x|y) \tag{4.2}$$

where we have chosen the constant K in the definition of the Shannon entropy

$$H(X) = -K \sum_x P(x) \ln P(x) \tag{4.3}$$

to be $K = 1/\ln 2$. With this choice of K, entropy is said to be measured in *bits*. The conditional entropy is a measure of the average uncertainty that remains about x when y is known. The mutual information $I(X;Y)$ for continuous probability densities is given by equation 12 in the Marzen, et al. paper. This expression can be derived from the definition of the mutual information, which is

$$I(X;Y) \equiv H(X) - H(X|Y). \tag{4.4}$$

Describe in words what the mutual information measures. What does $I(N_{\text{open}}; c)$ mean specifically for the case of MWC molecules as sensors of ligand concentration?

c) Derive equation 12 in the Marzen, et al. paper from the definition of $I(X;Y)$, equation (4.4). First use discrete probability distributions to show that

$$I(X;Y) = \sum_x \sum_y P(x) P(y|x) \log_2 \frac{P(y|x)}{P(y)}. \tag{4.5}$$

The continuous limit is then trivially taken.

d) When doing part(c), why can you not start with entropy for a continuous probability density? I.e., what is wrong with the expression below (written in red to indicate that it has a problem)?

$$H(X) = -K \int dx\, P(x) \ln P(x). \tag{4.6}$$

e) Describe in words what the channel capacity, I_{opt}, describes about the ion channel system. (Note: the word "channel" in channel capacity is a generic term; it does not specifically refer to ion channels.) Comment on equation 24 in the Marzen, et al. paper and its associated plots in Figure 10(d) and 10(e). In particular, what "knobs" can evolution turn to give a cell an efficient ligand concentration detection system[1]? If you were designing a cell, to what values would you tune these knobs?

Problem 4.3 (Measuring copy numbers, based on problem 2.7 of *PBoC2*, 15 pts)**.**
We have been studying gene expression lately in lecture, in the readings, and also in this homework. Experimentally, measurement of gene expression can be a tricky business. Expression is often read out by fluorescent intensity of a fluorescently tagged gene of interest. This is useful for computing fold change of expression, but it is not immediately obvious how to compute total numbers of expressed proteins.

A clever method to compute absolute numbers of fluorescent proteins was worked out in Rosenfeld, et al., *Science*, **307**, 1962–1965, 2005. They noted than in a colony of bacterial cells, the average square fluorescent intensity difference between two daughter cells is related to the total fluorescent intensity of the mother cell by

$$\langle (I_1 - I_2)^2 \rangle = \alpha I_{\text{tot}}. \tag{4.7}$$

Here, α is the fluorescent intensity of a single protein such that $I_1 = \alpha N_1$, where N_1 is the number of fluorescent proteins in daughter cell 1. The total fluorescent intensity of the mother cell immediately before division is $I_{\text{tot}} = I_1 + I_2$.

a) Derive equation (4.7). *Hint*: Think about the probability distribution that describes the inheritance of fluorescent proteins for each daughter cell.

b) Comment on what Figure 1 says about the validity and applicability of equation (4.7).

[1] By "knob," I mean physical parameters that can be changed to affect I_{opt}. The word is used in a similar context in descriptions of repression of gene expression in the Boedicker, et al. paper in your assigned reading.

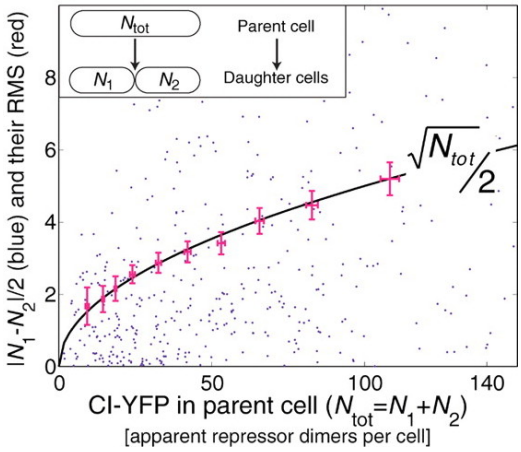

Figure 1: Figure modified from Rosenfeld, et al., *Science*, **307**, 1962–1965, 2005. The authors inferred values of N_1 and N_2 using a value of α obtained from performing a curve fit using equation (4.7). "RMS" in the y-axis label refers to "root mean square." In other words, the y-axis is $\sqrt{\langle (N_1 - N_2)^2 \rangle}/2$.

Problem 4.4 (States and weights for a repressor in a genetic switch, 10 pts)**.**
In lecture, we considered a genetic switch where protein 1 represses expression of protein 2 and protein 2 represses expression of protein 1. We assumed each repressor had the same physical constants with respect to binding to operators. Considering one of the repressors binding to its operon, we treated the case where either zero, one, or two repressors could be bound to an operator. There was no cooperativity; each binding event to an operator has the same ΔE_{rd}. We wrote that the fold change in gene expression due to this repression is

$$f_2(p_1) = \frac{1}{(1 + Kp_1)^2}, \tag{4.8}$$

where p_1 is the copy number of protein 1. Derive this result. In particular, write down an expression for K.

Problem 4.5 (How to make a genetic switch, 10 pts)**.**
Consider a genetic circuit as in problem 4.4 in which protein 1 represses expression of protein 2 and protein 2 represses expression of protein 1. The repression mechanism for each is "simple repression," as we defined in lecture to be the case where a single protein molecule binds to an operator. Show that this system cannot function as a genetic switch.

Problem 4.6 (A simplified repressilator, 30 pts)**.**
In this problem, we study a synthetic genetic circuit developed by Michael Elowitz and Stan Leibler called the repressilator. It is described in Elowitz and Leibler, *Nature*, **403**, 335–338, 2000. The circuit consists of three genes, lacI, tetR, and cI, that repress each other in a cyclic fashion. Another gene with a tet-repressible promoter was fused to green fluorescent protein (GFP) for a readout. So, if tetR has low copy numbers, we will see a large GFP signal and vice versa. A diagram of the repressive interactions of the genes is shown in Figure 2. So notation does not get cumbersome, we will refer to lacI as "1", tetR as "2", and cI as "3". The copy number of protein i per cell is p_i.

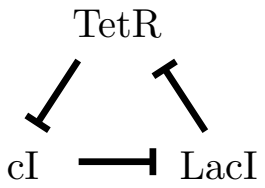

Figure 2: Schematic of the repressilator described in Elowitz and Leibler, *Nature*, **403**, 335–338, 2000.

a) Give an intuitive explanation as to why the repressilator system can give rise to oscillatory gene expression.

b) Write down a system of ODEs describing the time evolution of the p_i's. Ignore mRNA dynamics. That is, write down ODEs similar to those we wrote in class for the synthetic genetic switch. For the purposes of this problem, assume that the fold change in expression of a repressed gene is given by a Hill function with Hill coefficient n. (We explored when this assumption is valid in homework 3.)

c) Nondimensionalize these equations. As a simplifying assumption, take all phenomenological coefficients of each protein to be the same. I.e., they all have the same degradation rate, they all have the same basal production rate, etc. Your result should be of the form

$$\frac{\mathrm{d}p_1}{\mathrm{d}t} = -p_1 + \frac{\alpha}{1+p_3^n} \tag{4.9}$$

$$\frac{\mathrm{d}p_2}{\mathrm{d}t} = -p_2 + \frac{\alpha}{1+p_1^n} \tag{4.10}$$

$$\frac{\mathrm{d}p_3}{\mathrm{d}t} = -p_3 + \frac{\alpha}{1+p_2^n}, \tag{4.11}$$

where p_i now has a constant absorbed into it.

d) Show that this system has a unique fixed point.

e) Use linear stability analysis to show derive the stability properties of the fixed point. Specifically, show that

$$\text{the fixed point is } \begin{cases} \text{stable} & \text{for all } \alpha \text{ if } n \leq 2 \\ \text{stable} & \text{if } n > 2 \text{ and } \alpha < \dfrac{n}{2}\left(\dfrac{n}{2}-1\right)^{-\frac{n+1}{n}} \\ \text{unstable} & \text{if } n > 2 \text{ and } \alpha > \dfrac{n}{2}\left(\dfrac{n}{2}-1\right)^{-\frac{n+1}{n}}. \end{cases} \tag{4.12}$$

From this result, what can you say about the role of cooperativity in the repressilator system? *Hint*: In doing the linear stability analysis, it will help you to recall that there are three cube roots of unity.

$$\sqrt[3]{1} = \left\{1, -\frac{1}{2}\left(1+i\sqrt{3}\right), -\frac{1}{2}\left(1-i\sqrt{3}\right)\right\}. \tag{4.13}$$

You may also want to read the brief primer on linear stability analysis posted on the course website here: http://beaph161.caltech.edu/2014/handouts/linear_stability_analysis.pdf

f) Solve the repressilator system numerically for $n = 3$ and $\alpha = 3$, $\alpha = 10$, and $\alpha = 100$. Plot and comment on your results.